A Reproducible Research Framework for Audio Inpainting

Amir Adler and Michael Elad
Computer Science Departement - The Technion
Haifa 32000, Israel
E-mail: {adleram,elad}@cs.technion.ac.il

Valentin Emiya and Rémi Gribonval
INRIA Rennes - Bretagne Atlantique
35042 Rennes Cedex, France
Email: firstname.lastname@inria.fr

Maria G. Jafari and Mark D. Plumbley
Queen Mary University of London
London E1 4NS, U.K.
Email: firstname.lastname@elec.qmul.ac.uk

***Abstract*—We introduce a unified framework for the restoration of distorted audio data, leveraging the Image Inpainting concept and covering existing audio applications. In this framework, termed Audio Inpainting, the distorted data is considered missing and its location is assumed to be known. We further introduce baseline approaches based on sparse representations.**

For this new audio inpainting concept, we provide reproducible-research tools including: the handling of audio inpainting tasks as inverse problems, embedded in a frame-based scheme similar to patch-based image processing; several experimental settings; speech and music material; OMP-like algorithms, with two dictionaries, for general audio inpainting or specifically-enhanced declipping.

I. Introduction

Inpainting is a task proposed in the field of image processing: a set of missing pixels is reconstructed from the other reliable pixels of the image. Inpainting can be generalized as a problem of missing data estimation and techniques for image inpainting can be adapted to inpainting of other kinds of signals: one observes a partial set of reliable data while the remaining unreliable data is considered missing and is estimated from the reliable data. In particular, we consider Audio Inpainting [1] as a general task that covers a family of audio applications, including click removal, declipping, packet loss concealment and several applications for the restoration of time-frequency coefficients. We present works for audio inpainting in the time-domain [1], [2] and provide contributions on how to process audio signals in this context, which applicative scenarios and benchmarks are worth addressing and how sparse representations can solve those problems efficiently.

II. Audio Inpainting in time domain

A. Global and local formulation of Audio Inpainting

Let us consider a vector $\mathbf{s} \in \mathbb{R}^L$ of audio data. We only observe a subset of reliable samples $\mathbf{y}^{\mathrm{r}} = \mathbf{M}^{\mathrm{r}}\mathbf{s}$, where $\mathbf{y}^{\mathrm{r}} \in \mathbb{R}^{L'}$, $L' < L$ and $\mathbf{M}^{\mathrm{r}}$ is the so-called measurement matrix obtained from the $L \times L$ identity matrix by selecting the rows associated with the observed reliable coefficients in $\mathbf{s}$. The audio inpainting problem is defined as the recovery of the original signal $\mathbf{s}$ based on the knowledge of:

1) the reliable data $\mathbf{y}^{\mathrm{r}}$,
2) the support of the missing data (or, equivalently, $\mathbf{M}^{\mathrm{r}}$),
3) additional information about the observed signal,
4) and, optionaly, information about the missing data (e.g. in the case of clipping below).

As in many audio processing tasks and similarly to patch-based image processing, the signal can be locally modeled and processed: it is segmented into frames; each frame is then inpainted; the full restored signal is finally synthesized using an overlap-add method. Thus, the above global formulation of the inpainting problem can be straightforwardly translated locally at the frame level.

B. Audio Inpainting Problems

We propose **several scenarios or Problems** in which new inpainting algorithms can be compared against existing ones. They are related to speech or music restoration in different applications.

1) Isolate-sample-to-large-hole Problem: audio signals are degraded by periodically removing N_{miss} samples and performance are assessed as a function of N_{miss}. Small values of N_{miss} represent the click removal problem while large values of N_{miss} are simulating the packet loss concealment problem.

2) Missing-sample-topology Problem: for a fixed number of missing samples N_{miss} in a frame, a segments of b consecutive missing samples must be inpainted, where $a \times b = N_{\text{miss}}$. The performance is then reported as a function of the hole size b.

3) Declipping Problem: the missing samples are those beyond the clipping level θ_{clip}, such that the observation at time t is $\mathbf{y}^{\mathrm{r}}(t) = \mathbf{s}(t)$ if $|\mathbf{s}(t)| < \theta_{\text{clip}}$, $\mathbf{y}^{\mathrm{r}}(t) = \operatorname{sign}(\mathbf{s}(t))\,\theta_{\text{clip}}$ otherwise.

III. Baseline Dictionaries and Solvers

We propose sparsity-based approaches to address the Audio Inpainting problems described in Section II-B. **Two dictionaries** known to provide good models for audio waveforms are used: a discrete cosine transform dictionary, where phases are locked, and a free-phase Gabor dictionary. As a Solver, the inpainting version of the **OMP algorithm** is used to inpaint audio frames. We propose an enhancement for audio declipping, where the missing samples are constrained to have an amplitude beyond the clipping level.

IV. Material for reproducible research

For reproducible-research purposes, we provided GPL Matlab code and Creative Commons data related to the presented works and arranged in a Problems/Dictionaries/Solvers architecture as in [3]:

- a series of Problems described in Section II-B, including experiment generation, result display and speech and music datasets;
- an analysis/synthesis scheme to address the Problems by just inserting any frame-level inpainting solver (see Section II-A);
- the Dictionaries and Solvers proposed in Section III.

Acknowledgment

This work was supported by the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL: Sparse Models, Algorithms and Learning for Large-Scale data.

References

[1] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio Inpainting," *submitted to IEEE Trans. Audio, Speech, and Lang. Proc.*, 2011, http://hal.inria.fr/inria-00577079/en/.

[2] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A Constrained Matching Pursuit Approach to Audio Declipping," in *Proc. of ICASSP*, Prague, Czech Republic, May 2011.

[3] I. Damnjanovic, M. Davies, and M. D. Plumbley, "SMALLbox - an evaluation framework for sparse representations and dictionary learning algorithms," in *Proc. of LVA/ICA*, 2010, pp. 418–425.